Turkish Online Journal of Qualitative Inquiry (TOJQI)
Volume 11, Issue 3, October 2020: 625-643

A Second Order Smoothing Penalty Function Algorithm for Constrained Optimization Problems

Darpan Sood [1], Dr. Amanpreet Singh [2], Dr.Rama [3], Dr. Amrit Pal Singh [4]

1Research Scholar, Department of Mathematics, Desh Bhagat University, Punjab, (India).147301

2Assistant Professor, Department of Mathematics, GSSDGS Khalsa College, Patiala. Punjab, 147001

3Professor, Department of Mathematics, Desh Bhagat University, Punjab, (India).147301

4Assistant Professor, Department of Mathematics, SMHS Government College , SAS Nagar ,Punjab, 160055

Abstract The current paper introduces a second order smoothing technique for classical l_1 exact penalty function in constrained optimization problems. Error calculations for optimum solution values for non-smoothed, smoothed penalty problem and for the original problem have been discussed in the paper. An algorithmic procedure for obtaining the solution is demonstrated and convergence is discussed.

Keywords Penalty Function, Smoothing, Error , Convergence, Constrained optimization problem

Introduction

The mathematical form of constraint optimization problem involves the introduction of certain terminology which should be known for better understanding of the topic. Let x be an n-dimensional vector given as $x = (x^1, x^2, ...x^n) \in \mathrm{R}^n$, S be a subset of R^n. Let $f_0(x), f_1(x),..., f_n(x)$ are functions of x. The main problem in constrained optimization can be represented as

(1) $$Min \ f_0(x)$$

s.t $f_j(x) \le 0 \ \ j = 1,2,3,..., m$

The function f_0 and $f_j : R^n \to R$ are continuous differentiable functions of second order. The function $f_0(x)$ is called the objective function. The vector function $f(x) = f_1(x), f_2(x),..., f_m(x)$ defined above is generally referred to as the functional constraints. The set S is called the basic feasible set. The set $Q = \{x \in S, \ f_j(x) \le 0, \ j = 1,2,3,..., m\}$ is called the feasible set of the problem (1). The set Q is assumed to be nonempty. The minimization problems can be classified as

1. Constrained Problems: $Q \subset R^n$

2. Unconstrained Problems: $Q \equiv R^n$.
3. Smooth Problems: All the functions $f_j(x)$ are derivable in nature.
4. Non Smooth Problems: There is a function in the vector function which is not differentiable.
5. Linear Constrained Problems: All the functional constraints are linear in nature. The function $f_0(x)$ defined in (1) if is also linear then the problem in (1) is called linear programming problem. If the function $f_0(x)$ is quadratic then the problem is called Quadratic Programming Problem.

On the basis and properties of the feasibleset,the problem defined in (1) can also be classified as

a) The problem in (1) is called feasible if $Q \neq \phi$.
b) The problem stated in (1) is called strictly feasible if $for\, x \in Q \quad f_j(x) < 0$ (or > 0) for all the inequality constraints and $f_j(x) = 0$ for all equality constraints.

The solutions to the problems defined in (1) can also be classified as

a. x^* is called the global optimal solution if $f_0(x^*) \leq f_0(x)$ for all $x \in Q$. The solution value $f_0(x^*)$ is called the global optimum value.
b. x^* is called the local solution if $f_0(x^*) \leq f_0(x)$ for all $x \in \text{int}\, \overline{Q} \subset Q$.

Let us consider some examples which can explain the origin of constrained problems:

Consider the problem as

(2) Find $x \in R^n$ such that

$$f_1(x) = a_1$$
$$f_2(x) = a_2$$

$$f_m(x) = a_m$$

Now we can define a problem as $\min \sum_{j=1}^{m} (f_j(x) - a_j)^2$ with some additional constraints. This is a global problem and can be used in various fields of mathematics comprising of partial differential equations, ordinary differential equations, game theory etc.

Sometimes due to various constraints the decision variables are treated as integers giving rise to integer programming problems which is defined in (3) as

(3) $$Min\ f_0(x)$$

$$\text{s.t } a_j \le f_j(x) \le b_j\ ,\ \text{j} = 1,2,3...,\text{m}$$

$$x \in S$$

The problem defined in (1) can be found in the field of Business Management, Network Structure and Flow Problems, Traffic Control Problems and other large number of areas. Large amount of work has been done in past to solve the problem proposed in (1) by various authors involving the use of various methods. The penalty function method is also an important method developed and evolved in recent past to solve (1). The principle behind the concept of penalty function is to construct a sequence of unconstrained optimization problems in regard to (1) which can be easily solved by various methods and involves less labour. Recent past is replete of work done by various authors in the field of penalty function [1-5]. The first approach to solve (1) using penalty function was introduced by Zangwill [1] in which he introduced the notion of classical penalty function:

(4) $$U_1(x,\sigma) = f_0(x) + \sigma \sum_{j \in I} \max\{f_j(x),0\}$$

The equation (4) can also be written as $U_1(x,\sigma) = f_0(x) + \sigma \sum_{j=1}^{m} p(f_j(x))$

The corresponding penalty optimization problem using (4) for (1) is defined as

(5) $$Min U_1(x,\sigma) \quad \text{s.t} \quad x \in R^n$$

Another famous penalty function called the square, twice or l_2 penalty function which is defined as

(6) $$U_2(x,\sigma) = f_0(x) + \sigma \sum_{j \in I} \left[\max\{f_j(x),0\}\right]^2 .$$

The penalty parameter σ is >0. The function is continuously differentiable or smooth but is not an exact penalty function. The apparent problem with the penalty function defined in (4) is its non - differentiable nature which prevents the use of algorithms based on derivability and thus increasing instability. Huang and Yang et al. [2-5] have studied a new type of penalty function defined as

(7) $$U_k(x,\sigma) = \left[(f_0(x))^k + \sigma \sum_{j \in I} \max\{f_j(x),0\}^k \right]^{\frac{1}{k}}$$

and is known as k order penalty function and also being studied in [6-7]. The penalty function defined in (7) becomes the classical l_1 penalty function if the value of k is taken as 1. The penalty function is smooth for $k > 1$ and is not smooth for $0 < k \le 1$. Another lower order penalty function is discussed by Meng et al. [8] and Wu et al. [9] of the form

(8) $$U^k(x,\sigma) = f_0(x) + \sigma \sum_{j \in I} [\max\{f_j(x), 0\}]^k$$

here also the function is not derivable for $k \in (0,1)$.

Meng et al. [10] and Wu et al. [11] have also proposed an exact penalty given in (9) and its smoothing techniques to solve the problem.

(9) $$U_{\frac{1}{2}}(x,\sigma) = f_0(x) + \sigma \sum_{j \in I} \left[\max\{f_j(x), 0\}\right]^{\frac{1}{2}}$$

The smoothing of penalty functions is of paramount importance as it plays a magnificent role in solving (1) using various methods which are more or less based on differentiable functions. Various authors in [6-9] and [12-16] have proposed smoothing techniques for classical penalty function for constrained non -linear programming problems. Yang et al. [6-7] have demonstrated smoothing techniques for (9). Many of the methods that are used to solve unconstrained optimization problems involves the use of second order derivatives so it has become imperative to form a smoothing technique which is second order differentiable to the exact penalty function. In this chapter we will try to formulate a smoothing technique for classical penalty function. In this chapter we will try to formulate a smoothing technique for classical l_1 penalty function in terms of second order differentiability. Let us define a function

(10) $$p_\varepsilon(t) = \begin{cases} 0, t < 0 \\ \dfrac{t^4}{10\varepsilon^3}, 0 \le t < \varepsilon \\ t - \dfrac{3\varepsilon}{2} + \dfrac{3\varepsilon^2}{5t}, \text{t} \ge \varepsilon \end{cases}$$

Using the function defined above a second order differential approximation to the classical exact penalty function will be developed in this chapter and an algorithm hinging on the smoothing technique will be developed for solving the constrained optimization problem. Consider the function

(11) $p(t) = \max\{t, o\}$

Using (11) the problem in (5) becomes

(12) $Min U_1(x,\sigma) = f_0(x) + \sigma \sum_{j \in I} p(f_j(x))$ such that $x \in R^n$.

For $\sigma > 0$, let

$$p_\varepsilon(t) = \begin{cases} 0, t < 0 \\ \frac{t^4}{10\varepsilon^3}, 0 \le t < \varepsilon \\ t - \frac{3\varepsilon}{2} + \frac{3\varepsilon^2}{5t}, t \ge \varepsilon \end{cases}$$

where $\varepsilon > 0$ is called smoothing parameter. The function $p_\varepsilon(t)$ has some abstractive properties which are proved in the theorem given below.

Theorem 1

Statement: For any $\varepsilon > 0$ we have the following results

(i) $p_\varepsilon(t)$ is twice continuously differentiable in t for all $\varepsilon > 0$.

(ii) for each t in $R, p(t) \ge p_\varepsilon(t)$

(iii) $lt_{\varepsilon \to 0} p_\varepsilon(t) = p(t)$

Proof:

(i)

The second order continuous differentiability can be evaluated by checking the derivability at 0 and ε.

Continuity at 0

LHL= $lt_{t \to 0^-} p_\varepsilon(t) = lt_{t \to 0^-} 0 = 0$

RHL= $lt_{t \to 0^+} p_\varepsilon(t) = lt_{t \to 0^+} \frac{t^4}{10\varepsilon^3} = lt_{h \to 0} \frac{(0+h)^4}{10\varepsilon^3} = 0$

Hence LHL=RHL, Thus the function is continuous at 0

Now we will discuss the continuity of the function at ε.

LHL= $lt_{t \to \varepsilon^-} p_\varepsilon(t) = lt_{t \to \varepsilon^-} \frac{t^4}{10\varepsilon^3} = lt_{h \to 0} \frac{(\varepsilon - h)^4}{10\varepsilon^3} = \frac{\varepsilon}{10}$

RHL= $lt_{t \to \varepsilon^+} p_\varepsilon(t) = lt_{t \to \varepsilon^+} t - \frac{3\varepsilon}{2} + \frac{3\varepsilon^2}{5t} = lt_{h \to 0} (\varepsilon + h) - \frac{3\varepsilon}{2} + \frac{3\varepsilon^2}{5(\varepsilon + h)} = \varepsilon - \frac{3\varepsilon}{2} + \frac{3\varepsilon}{5} = \frac{\varepsilon}{10}$

Hence RHL=RHL, thus proving that the function is continuous at ε. The above two results prove the continuity of $p_\varepsilon(t)$ at 0 and ε. The first derivative of the function $p_\varepsilon(t)$ is given as

$$p_\varepsilon{}'(t) = \begin{cases} 0, t < 0 \\ \dfrac{4t^3}{10\varepsilon^3}, 0 \le t < \varepsilon \\ 1 - \dfrac{3\varepsilon^2}{5t^2}, \text{t} \ge \varepsilon \end{cases}$$

Now we are going to prove the continuity of first order derivative at 0 and ε which will establish the first order continuous differentiability of the function.

Continuity at t=0

LHL= $lt_{t\to 0^-} 0 = 0$

RHL= $lt_{t\to 0^+} p_\varepsilon{}'(t) = lt_{t\to 0^+} \dfrac{4t^3}{10\varepsilon^3} = lt_{h\to 0} \dfrac{4(0+h)^3}{10\varepsilon^3} = 0$

Hence LHL =RHL , Thus the function is continuous at 0.

Continuity at t= ε

LHL= $lt_{t\to \varepsilon^-} p_\varepsilon{}'(t) = lt_{t\to \varepsilon^-} \dfrac{4t^3}{10\varepsilon^3} = lt_{h\to 0} \frac{4(\varepsilon-h)^3}{10\varepsilon^3} = \dfrac{4}{10} = \dfrac{2}{5}$

RHL= $lt_{t\to \varepsilon^+} p_\varepsilon{}'(t) = lt_{t\to \varepsilon^+} 1 - \dfrac{3\varepsilon^2}{5t^2} = lt_{h\to 0} 1 - \dfrac{3\varepsilon^2}{5(\varepsilon+h)^2} = 1 - \dfrac{3}{5} = \dfrac{2}{5}$

Hence LHL=RHL, thus the function is continuous at ε. Hence the derivative of the function is continuous thus establishing the fact that function is first order continuous differentiable. Let us move further to prove the second order continuous differentiability of the function. The second order derivative of the smoothing function is given as

$$p_\varepsilon{}''(t) = \begin{cases} 0, t < 0 \\ \dfrac{12t^2}{10\varepsilon^3}, 0 \le t < \varepsilon \\ \dfrac{6\varepsilon^2}{5t^3}, \text{t} \ge \varepsilon \end{cases}$$

Let us now check the continuity of the above defined second order derivative of the smoothing function at 0 and ε.

Continuity at t=0

LHL= $lt_{t\to 0^-} p_\varepsilon''(t) = lt_{t\to 0^-} 0 = 0$

RHL= $lt_{t\to 0^+} p_\varepsilon''(t) = lt_{t\to 0^+} \frac{12t^2}{10\varepsilon^3} = lt_{h\to 0} \frac{12(0+h)^2}{10\varepsilon^3} = 0$

Hence LHL=RHL, thus the second order derivative is continuous at t=0.

Continuity at t= ε

LHL= $lt_{t\to \varepsilon^-} p_\varepsilon''(t) = lt_{t\to \varepsilon^-} \frac{12t^2}{10\varepsilon^3} = lt_{h\to 0} \frac{12(\varepsilon-h)^2}{10\varepsilon^3} = \frac{6}{5\varepsilon}$

RHL= $lt_{t\to \varepsilon^+} p_\varepsilon''(t) = lt_{t\to \varepsilon^+} \frac{6\varepsilon^2}{5t^3} = lt_{h\to 0} \frac{6(\varepsilon)^2}{5(\varepsilon+h)^3} = \frac{6}{5\varepsilon}$

Hence the function is continuous at t= ε .

Hence the function is second order continuous differentiable. This completes the proof of (i).

(iii)

Consider the function defined in (10) and (11) as

$$p(t) = \begin{cases} t, t \geq 0 \\ 0, t < 0 \end{cases}$$

and

$$p_\varepsilon(t) = \begin{cases} 0, t < 0 \\ \frac{t^4}{10\varepsilon^3}, 0 \leq t < \varepsilon \\ t - \frac{3\varepsilon}{2} + \frac{3\varepsilon^2}{5t}, \text{t} \geq \varepsilon \end{cases}$$

Now consider the difference

$$p(t) - p_\varepsilon(t) = \begin{cases} 0, t < 0 \\ t - \frac{t^4}{10\varepsilon^3}, 0 \leq t < \varepsilon \\ \frac{3\varepsilon}{2} - \frac{3\varepsilon^2}{5t}, \ t \geq \varepsilon \end{cases}$$

When $0 \leq t < \varepsilon$,

Define

$$Z(t) = t - \frac{t^4}{10\varepsilon^3}$$

$$Z'(t) = 1 - \frac{4t^3}{10\varepsilon^3}$$

$$\Rightarrow Z'(t) > 0$$

Now $Z(0) = 0$ and $Z(\varepsilon) = \varepsilon - \frac{\varepsilon^4}{10\varepsilon^3} = \varepsilon - \frac{\varepsilon}{10} = \frac{9\varepsilon}{10}$

$$\Rightarrow \quad 0 \le p(t) - p_\varepsilon(t) \le \frac{9\varepsilon}{10}$$

Hence $lt_{t\to 0}\, p_\varepsilon(t) = p(t)$.

(ii)

From the result proved above it is quite apparent that

$$p(t) - p_\varepsilon(t) \ge 0$$

Hence $p(t) \ge p_\varepsilon(t)$

Hence $p_\varepsilon(t)$ is monotonically increasing in t for any $\varepsilon > 0$.

Hence Proved.

Also we see that $p(f_j(x) = \max\{0, f_j(x)\}$

Supposing that f_0 and f_j ($j = 1,2,3,..,m$) are continuous differentiable functions of second order and further let us state the penalty function

(13) $$U(x,\sigma,\varepsilon) = f_0(x) + \sigma \sum_{j \in I} p_\varepsilon(f_j(x))$$

where $\sigma > 0$ is the required penalty parameter. The function $U(x.\sigma,\varepsilon)$ is twice continuously differentiable at any point in R^n. The application of (13) will help us to formulate the corresponding penalty problem to (1) as

(14) $$Min U(x,\sigma,\varepsilon)$$

such that x lies in n-dimensional space R. Now we are in a position to study the functional relationship between the functions defined in (12) and (14).

Theorem 2

Statement:- Let for some $x \in R^n$ and $\varepsilon > 0$

$$0 \le U_1(x,\sigma) - U(x,\sigma,\varepsilon) \le \frac{9m\sigma\varepsilon}{10}$$

Proof:-

From the definition and the result proved in (i) of theorem , it can be said that

$$p(f_j(x)) - p_\varepsilon(f_j(x)) = \begin{cases} 0, \ f_j(x) < 0 \\ f_j(x) - \frac{f_j(x)^4}{10\varepsilon^3}, 0 \le f_j(x) < \varepsilon \\ \frac{3\varepsilon}{2} - \frac{3\varepsilon^2}{5f_j(x)}, \ f_j(x) \ge \varepsilon \end{cases}$$

As we have already proved above that

$$0 \le p(t) - p_\varepsilon(t) \le \frac{9\varepsilon}{10}$$

$$0 \le p(f_j(x) - p_\varepsilon(f_j(x)) \le \frac{9\varepsilon}{10}$$

$$0 \le \sigma(\sum_{j=1}^{m}(p(f_j(x) - p_\varepsilon(f_j(x))) \le \frac{9m\sigma\varepsilon}{10}$$

$$0 \le \sigma\sum_{j=1}^{m} p(f_j(x)) - \sigma\sum_{j=1}^{m} p_\varepsilon(f_j(x)) \le \frac{9m\sigma\varepsilon}{10}$$

$$0 \le \sigma[f_0(x) + \sum_{j=1}^{m} p(f_j(x))] - \sigma[f_0(x) + \sum_{j=1}^{m} p_\varepsilon(f_j(x))] \le \frac{9m\sigma\varepsilon}{10}$$

$$0 \le U_1(x,\sigma) - U(x,\sigma,\varepsilon) \le \frac{9m\sigma\varepsilon}{10}$$

Hence Proved.

The theorem proves that the gap between $U(x,\sigma,\varepsilon)$ and $U_1(x,\sigma)$ can be controlled by the help of the smoothing parameter. An important ramification of theorem 1 is proved in theorem 2.

Theorem 3

Statement:- For a sequence of positive numbers ε_j which approaches to 0 as $j \to \infty$, suppose that x_j is an optical solution to $\min_{x \in R^n} U(x, \sigma, \varepsilon_j)$ for some $\sigma > 0$.Let x' is accumulating point of the sequence $\{x_j\}$, then x' is an optical solution of the problem $\min_{x \in R^n} U_1(x, \sigma)$.

Proof:

From the result proved in the above theorem we have

(15) $$0 \le U_1(x, \sigma) - U(x, \sigma, \varepsilon_j) \le \frac{9m\sigma\varepsilon_j}{10}$$

Since x_j is an optical solution to $\min_{x \in R^n} U(x, \sigma, \varepsilon_j)$ gives us

(16) $U(x_j, \sigma, \varepsilon_j) \le U(x, \sigma, \varepsilon_j)$ for all $x \in R^n$

Now using the above two inequalities

$$U_1(x_j, \sigma) \le U(x_j, \sigma, \varepsilon_j) + \frac{9m\sigma\varepsilon_j}{10}$$

$$\le U(x, \sigma, \varepsilon_j) + \frac{9m\sigma\varepsilon_j}{10}$$

$$\le U_1(x, \sigma) + \frac{9m\sigma\varepsilon_j}{10}$$ (from (15) and rearranging)

As $j \to \infty$ leads to $\varepsilon_j \to 0$, apply it in the above inequality we get

$U_1(x', \sigma) \le U_1(x, \sigma)$ (since x' is accumulating point)

Hence x' is an optical solution to the $\min_{x \in R^n} U_1(x, \sigma)$.

Theorem 4

Statement:- Let x'' is an optimal solution of the problem defined in (5) or in (12). Suppose $x''' \in R^n$ is assumed to be optical solution of the problem defined in (14). Then we have the following result

$$0 \le U_1(x'', \sigma) - U(x''', \sigma, \varepsilon) \le \frac{9m\sigma\varepsilon}{10}$$

Proof:

Using the result obtained in theorem 2 and for the parameter $\sigma > 0$ we have

(17) $0 \le U_1(x'',\sigma) - U(x'',\sigma,\varepsilon) \le \frac{9m\sigma\varepsilon}{10}$

(18) $0 \le U_1(x''',\sigma) - U(x''',\sigma,\varepsilon) \le \frac{9m\sigma\varepsilon}{10}$

As x'' and x''' are assumed to be optical solution to the problem $\min U_1(x,\sigma)$ and $\min U(x,\sigma,\varepsilon)$ we get

(19) $U_1(x'',\sigma) \le U_1(x''',\sigma)$

(20) $U(x''',\sigma,\varepsilon) \le U(x'',\sigma,\varepsilon)$

Now using all the inequalities from (17), (18),(19),(20) we have

$0 \le U_1(x'',\sigma) - U(x'',\sigma,\varepsilon)$

$\le U_1(x'',\sigma) - U(x''',\sigma,\varepsilon)$

$\le U_1(x''',\sigma) - U(x''',\sigma,\varepsilon)$

$\le \frac{9m\sigma\varepsilon}{10}$

Thus we get

$0 \le U_1(x'',\sigma) - U(x''',\sigma,\varepsilon) \le \frac{9m\sigma\varepsilon}{10}$

Hence Proved.

Definition 1

Let x_ε be a point in n-dimensional space R. The point is called an ε-feasible solution or an ε solution if the following condition is satisfied that is

$f_j(x_\varepsilon) \le \varepsilon$, $j = 1,2,3,\ldots,m$

Let us now propose and prove an important theorem concerning the above definition.

Theorem 5

Statement: - Suppose $\bar{x}$ is an optimal solution of (12) and $\hat{x}$ be an optimal solution of (14) where both $\bar{x}$ and $\hat{x}$ lies in n-dimensional space R. Also further presume that $\bar{x}$ is a feasible solution to the problem defined in (1) and $\hat{x}$ is ε-feasible solution to the problem defined in (1), then

$\frac{m\sigma\varepsilon}{10} \le f_0(\bar{x}) - f_0(\hat{x}) \le m\varepsilon\sigma$

Proof:

Since $\hat{x}$ is presumed to be ε -feasible solution to (1)

(21) $\Rightarrow \sum_{j=1}^{m} p_{\varepsilon}(f_j(\hat{x}) \leq \frac{m\varepsilon}{10}$

Now it has been given in the statement that $\bar{x}$ is feasible solution to (1)

(22) $\Rightarrow \sum_{j=1}^{m} p(f_j(\bar{x}) = 0$

Thus using the result proved in theorem 4 we have

$$0 \leq U_1(\bar{x},\sigma) - U(\hat{x},\sigma,\varepsilon) \leq \frac{9m\sigma\varepsilon}{10}$$

$$0 \leq f_0(\bar{x}) + \sigma\sum_{j=1}^{m} p(f_j(\bar{x})) - f_0(\hat{x}) - \sigma\sum_{j=1}^{m} p_{\varepsilon}(f_j(\hat{x})) \leq \frac{9m\sigma\varepsilon}{10}$$

Using (21) and (22) in the above inequality we get

$$0 \leq f_0(\bar{x}) - \{f_0(\hat{x}) + \sigma\sum_{j=1}^{m} p_{\varepsilon}(f_j(\hat{x}))\} \leq \frac{9m\sigma\varepsilon}{10}$$

$$\sigma\sum_{j=1}^{m} p_{\varepsilon}(f_j(\hat{x})) \leq f_0(\bar{x}) - f_0(\hat{x}) \leq \frac{9m\sigma\varepsilon}{10} + \sigma\sum_{j=1}^{m} p_{\varepsilon}(f_j(\hat{x}))$$

$$\frac{m\sigma\varepsilon}{10} \leq f_0(\bar{x}) - f_0(\hat{x}) \leq \frac{9m\varepsilon}{10} + \frac{m\sigma\varepsilon}{10}$$

$$\frac{m\sigma\varepsilon}{10} \leq f_0(\bar{x}) - f_0(\hat{x}) \leq m\sigma\varepsilon$$

Thus the theorem is proved.

The above theorem shows that a near about optimal solution to (1) can can be obtained under some certain soft conditions by solving (12).

The functions defined in (1) can be convex functions also. A result describing the sub-optimality condition of an optimal solution to (1) bounded by the penalty and smoothing parameter can be established and is discussed in the theorem proved below. Before proceeding further to state and prove the theorem , let us define the notion of Lagrange’s multipliers for convex functions.

Definition 2

Forany $x^* \in R^n$, the vector $z^* \in R^m$ is called a Lagrangian augment vector corresponding to x^* iff x^* and z^* satisfy the necessary condition

(23) $$\nabla f_0(x^*) + \sum_{j=1}^{m} z_j^* \nabla f_j(x^*) = 0$$

(24) $$z_j^* f_j(x^*) = 0$$

(25) $$z_j^* \geq 0$$

(26) $$f_j(x^*) \leq 0 \text{ for } j = 1,2,3,\ldots\ldots,m$$

Theorem 6

Statement: - Suppose that the functions $f_0(x)$ and $f_j(x)$ defined in (1) are assumed to be convex functions. Let $\tilde{x}$ is presume to be an optimal solution of (1) and z^* be Lagrange Multiplier vector in reference to $\tilde{x}$ then

$$U(\tilde{x}, \sigma, \varepsilon) \leq U(x, \sigma, \varepsilon) + \frac{9m\varepsilon\sigma}{10}$$

provided that $\sigma > z_j^*$, $j = 1,2,3,\ldots,m$

Proof:

Using the idea of convexity of both the functions f_0 and f_j for $j = 1,2,3,\ldots,m$ we have the condition

(27) $f_o(x) \geq f_0(\tilde{x}) + \nabla f_0(\tilde{x})^T (x - \tilde{x})$ for all x in R^n also

(28) $f_j(x) \geq f_j(\tilde{x}) + \nabla f_j(\tilde{x})^T (x - \tilde{x})$, $j = 1,2,3,\ldots,m$

As $\tilde{x}$ is assumed to be optimal solution to (1) and z^* is supposed to be corresponding LaGrange vector multiplier, therefore by (23), (24), (25),(26), (27),(28) we can say that

$$U_1(x, \sigma) \geq f_0(\tilde{x}) + \nabla f_\circ(\tilde{x})^T (x - \tilde{x}) + \sigma \sum_{j=1}^{m} f_j^+(x)$$

$$= f_\circ(\tilde{x}) - \sum_{j=1}^{m} z_j^* \nabla f_j(\tilde{x})^T (x - \tilde{x}) + \sigma \sum_{j=1}^{m} f_j^+(x)$$

$$\geq f_\circ(\tilde{x}) - \sum_{j=1}^{m} z_j^* (f_j(x) - f_j(\tilde{x})) + \sigma \sum_{j=1}^{m} f_j^+(x)$$

$$= f_\circ(\tilde{x}) - \sum_{j=1}^{m} z_j^* f_j(x) + \sigma \sum_{j=1}^{m} f_j^+(x)$$

because $f_j(x) \leq f_j^+(x)$

hence we have

$$U_1(x,\sigma) \geq f_\circ(\tilde{x}) + \sum_{j=1}^{m} (\sigma - z_j^*) f_j^+(x)$$

So for $\sigma > z_j^*$ and for $j = 1,2,3,\ldots,m$ we get the following result

(29) $U_1(x,\sigma) \geq f_\circ(\tilde{x})$

Now from the theorem 2 we have the result

(30) $0 \leq U(x,\sigma,\varepsilon) - U_1(x,\sigma) \leq \dfrac{9m\varepsilon\sigma}{10}$

Since the result is true for every value of x so it must be true for $\tilde{x}$ also

(31) $\Rightarrow 0 \leq U(\tilde{x},\sigma,\varepsilon) - U_1(\tilde{x},\sigma) \leq \dfrac{9m\varepsilon\sigma}{10}$

Moreover $\tilde{x}$ is feasible solution so $f_\circ(\tilde{x}) = U_1(\tilde{x},\sigma)$ (using (12))

Putting the above equality in (31) we get

(32) $0 \leq U(\tilde{x},\sigma,\varepsilon) - f_\circ(\tilde{x}) \leq \dfrac{9m\sigma\varepsilon}{10}$

Also $\tilde{x}$ is feasible solution so we have

$$U_1(\tilde{x},\sigma) \leq U_1(x,\sigma)$$

(33) $U_1(\tilde{x},\sigma) + \dfrac{9m\varepsilon\sigma}{10} \leq U_1(x,\sigma) + \dfrac{9m\varepsilon\sigma}{10}$

Now from (31) we have the inequality

$$U(\tilde{x},\varepsilon,\sigma) \leq U_1(\tilde{x},\sigma) + \frac{9m\varepsilon\sigma}{10}$$

$$U(\tilde{x},\varepsilon,\sigma) - \frac{9m\varepsilon\sigma}{10} \leq U_1(\tilde{x},\sigma)$$

$$-U(\tilde{x},\varepsilon,\sigma) + \frac{9m\varepsilon\sigma}{10} \geq -U_1(\tilde{x},\sigma)$$

(34) $U(x,\varepsilon,\sigma) - U(\tilde{x},\varepsilon,\sigma) + \dfrac{9m\varepsilon\sigma}{10} \geq U(x,\varepsilon,\sigma) - U_1(x,\sigma)$

Also from the inequality (30) we can say that

$$U(x,\sigma,\varepsilon) - U_1(x,\sigma) \geq 0$$

Using this in the above inequality in (34)

$$U(x,\varepsilon,\sigma) - U(\tilde{x},\varepsilon,\sigma) + \frac{9m\varepsilon\sigma}{10} \geq 0$$

$$U(x,\varepsilon,\sigma) - U(\tilde{x},\varepsilon,\sigma) \geq -\frac{9m\varepsilon\sigma}{10}$$

$$-U(x,\varepsilon,\sigma) + U(\tilde{x},\varepsilon,\sigma) \leq \frac{9m\varepsilon\sigma}{10}$$

$U(\tilde{x},\varepsilon,\sigma) - U(x,\varepsilon,\sigma) \le \frac{9m\varepsilon\sigma}{10}$

$U(\tilde{x},\varepsilon,\sigma) \le U(x,\varepsilon,\sigma) + \frac{9m\varepsilon\sigma}{10}$

Hence Proved

The above result helps us to comprehend that when the value of penalty parameter is large enough than an appropriate solution to (12) is also an approximate solution to (1) if the functions are assumed to be convex.

Algorithm

On the basis of smoothing penalty function defined for smoothing penalty problem defined in (14) , we submit an algorithm to solve the problem defined in (1) in the following steps

Step 1: For given x°, $\varepsilon > 0$, $\varepsilon_{\circ} > 0$, $\sigma_{\circ} > 0$, $N > 1$ and $0 < \delta < 1$. Let $k = 0$ and move forward to step 2.

Step 2: Solve the problem $\min_{x \in R^n} U(x,\varepsilon_k,\sigma_k)$ taking x^k as the initial point. Let x^{k+1} is the solution obtained in the current step.

Step3: Check for ε feasibility of the solution for the problem (1). If it is found to be $\varepsilon -$ feasible then stop the procedure. If it is not so then evaluate $\sigma_{k+1} = N\sigma_k$, $\varepsilon_{k+1} = \delta\varepsilon_k$ and $k = k+1$ and proceed further as in step 2.

The step 3 of the above algorithm and by the theorem 5 it can be said that x^{k+1} is an approximate optimal solution to (1). Also it can be found that in the above algorithm that as the sequence ε_k approaches to the value 0 as $k \to \infty$ and $\sigma_k \to \infty$ for $k \to \infty$ and the $\{\varepsilon_k \sigma_k\}$ also decreases to 0.

Theorem 7

Statement:- Let $\arg\min_{x \in R^n} U(x,\varepsilon,\sigma) \ne \phi$ for any $\sigma \in [\sigma_{\circ},\infty), \varepsilon \in (0,\varepsilon_{\circ}]$. Further if the sequence $\{x^k\}$ developed by the algorithm defined above has an accumulation point, then any accumulation point of $\{x^k\}$ is the solution to (1).

Proof:

Let us assume that the accumulation point of the sequence generated is $\hat{x}$. Hence for some subset V of N we have $x^k \to \hat{x}$. In order to show that $\hat{x}$ is an optimal solution to the problem (1) it is sufficient to prove the following two conditions which are

a. $\hat{x} \in Q$
b. $f_{\circ}(\hat{x}) \le \inf_{x \in Q} f_{\circ}(x)$

a. let us assume that $\hat{x} \notin Q$ and and for some $\delta_{\circ} > 0$ there exists an index $j_{\circ}$ such that

(35) $f_{j_{\circ}}(x^j) \ge \delta_{\circ}$

Now from the third step in the above algorithm it is clear that $\sigma_k \to \infty$ as $k \to \infty$. Also from the second step of the algorithm it is evident that for any $x \in Q$ we have

$U(x^k,\varepsilon_k,\sigma_k) - U(x,\varepsilon_k,\sigma_k) \le \frac{9m\varepsilon_k\sigma_k}{10}$ (from theorem 6)

$\Rightarrow f_{\circ}(x^k) - f_{\circ}(x) \le \frac{9m\varepsilon_k\sigma_k}{10}$ which contradicts as $\sigma_k \to \infty$. Hence our assumption is false. Hence $\hat{x} \in Q$

b. From the algorithm we have the condition

$\varepsilon_k\sigma_k \to 0$ thus for any $\delta > 0$ we have $m\varepsilon_k\sigma_k \le \delta$. Also from the second step of the algorithm for any $x \in Q$

$$f_{\circ}(x^k) - f_{\circ}(x) \le \frac{9m\varepsilon_k\sigma_k}{10}$$

$$\le \frac{\delta}{10}$$

Thus $f_{\circ}(\hat{x})$ satisfies the defined condition.

Hence Proved.

Convergence

Define for all $x \in R^n$ sets

(36) $V_{\varepsilon}^{-}(x) = \{j \mid f_j(x) < \varepsilon, j = 1,2,3,\ldots,m\}$

(37) $V_{\varepsilon}^{+}(x) = \{j \mid f_j(x) \ge \varepsilon, j = 1,2,3,\ldots,m\}$

Now after defining the above two sets, let us now discuss the convergence in the following theorem.

Theorem 8

Statement:- Let the function satisfies the coercive condition which is $lt_{\|x\|\to\infty} f_{\circ}(x) = +\infty$ and $\{x^k\}$ be the sequence generated in the algorithm. Assume that $U(x^k, \varepsilon_k, \sigma_k)$ is bounded sequence. Then prove that the sequence $\{x^k\}$ is also a bounded sequence and every accumulation point $\breve{x}$ is an optimal solution to the problem defined in (1).

Proof:

The proof of the above theorem is divided into two parts. The first part is for proving the boundedness of the sequence $\{x^k\}$ and in the second half is to prove that every accumulation point is also an optimal solution to (1).

Let us now prove that the sequence is bounded. Since it is already presumed in the statement that the sequence

$U(x^k, \varepsilon_k, \sigma_k)$ is bounded , hence there exists a no M such that

(38) $U(x^k, \varepsilon_k, \sigma_k) \le M \quad j = 0,1,2,3,\ldots.$

Now assume that the sequence $\{x^k\}$ is unbounded and wlog $\|x^k\| \to \infty$ as $k \to \infty$.

From the definition of smoothing function it is known that

(39) $$\sum_{j \in I} p_{\varepsilon}(f_j(x^k) \ge 0$$

Therefore from (38) and (39)

$$M \geq U(x^k, \varepsilon_k, \sigma_k) = f_\circ(x^k) + \sigma_j \sum_{j \in I} p_{\varepsilon_k}(f_j(x^k)) \geq f_\circ(x^k), \qquad j = 0,1,2,\ldots$$

which is contrary to what have been supposed in (38). Hence the sequence $\{x^k\}$is bounded.

Further it would be shown that $\breve{x}$ is feasible solution to (1). Suppose that

(40) $lt_{k \to \infty} x^k = \breve{x}$.

Assume that $\breve{x}$ is not feasible to (1). Therefore there exists some $j \in I$ such that $f_j(\breve{x}) \geq \mu > 0$. Now

(41) $U(x^k, \varepsilon_k, \sigma_k) = f_\circ(x^k) + \sigma_k \sum_{j \in I} p_{\varepsilon_k}(f_j(x^k))$

$$= f_\circ(x^k) + \sigma_k \sum_{j \in V^+_{\varepsilon_k}(x^k)} p_{\varepsilon_k}(f_j(x^k)) + \sigma_k \sum_{j \in V^-_{\varepsilon_k}(x^k)} p_{\varepsilon_k}(f_j(x^k))$$

$$= f_\circ(x^k) + \sigma_k \sum_{j \in V^+_{\varepsilon_k}(x^k)} f_j(x^k) + \frac{3\varepsilon_k^2}{5 f_j(x^k)} - \frac{3\varepsilon_k}{2} + \sigma_k \sum_{j \in V^-_{\varepsilon_k}(x^k)} \frac{f_j(x^k)^4}{10\varepsilon_k^3}$$

Since $f_j(\breve{x}) \geq \mu > 0$, so for any large k the set $j : f_j(\breve{x}) \geq \mu$ is non -void . Then any $j' \in I$ will satisfy the condition $f_{j'}(\breve{x}) \geq \mu$. So if $k \to \infty$ and $\sigma_k \to \infty$ and $\varepsilon_k \to 0$, then (41) gives

$U(x^k, \varepsilon_k, \sigma_k) \to \infty$ which contradictory to what has been assumed in the statement of the theorem. Hence our assumption is false. Thus $\breve{x}$ is a feasible solution to (1). Now in the last part of the proof it is required to show that $\breve{x}$ is an optical solution to (1). Let us suppose that $\bar{x}$ is an optical solution to (1). Since in the algorithm it has been assumed that x^k is an optical solution to $U(x^k, \varepsilon_k, \sigma_k)$ which means that

(42) $U(x^k, \varepsilon_k, \sigma_k) \leq U(\bar{x}, \varepsilon_k, \sigma_k) \quad k = 1.2.3....$, gives

$$f_\circ(x^k) + \sigma_k \sum_{j \in I} p_{\varepsilon_k}(f_j(x^k) \leq f_\circ(\bar{x}) + \sigma_k \sum_{j \in I} p_{\varepsilon_k}(f_j(\bar{x}))$$

Now using (39) in above inequality it is seen that

(43) $f_\circ(x^k) + \sigma_k \sum_{j \in I} p_{\varepsilon_k}(f_j(x^k) \leq f_\circ(\bar{x})$

Let $k \to \infty$, using (40) in (43)

(44) $f_\circ(\breve{x}) + \sigma_k \sum_{j \in I} p_{\varepsilon_k}(f_j(\breve{x}) \leq f_\circ(\bar{x})$

as $\breve{x}$ is feasible to (1) , it can be concluded that

(45) $f_\circ(\breve{x}) \leq f_\circ(\bar{x})$

Also $\breve{x}$ is feasible to (1) and $\bar{x}$ is an optical solution to (1) , it gets us the inequality

(46) $f_\circ(\bar{x}) \leq f_\circ(\breve{x})$

So from (45) and (46)

$f_{\circ}(\bar{x}) = f_{\circ}(\breve{x})$

Thus $\breve{x}$ is an optimal solution to (1).

References

1. Zangwill, W.I , Non Linear Programming via Penalty Functions. Management Science 13(5), 334-358(1967).

2 Huang, X.X, Yang.X.Q, Duality and exact penalization via augmented Lagrangian. J.Optim. Theory Appl. 111(3), 615-640(2001)

3 Rubinov, A.M. , Glover,B.M. , Yang, X.Q. , Extended Lagrange and penalty functions in continuous optimization 46(4) , 327-351(1999).

4 Rubinov, A.M. , Glover,B.M. , Yang, X.Q. ,Decreasing functions with applications to penalization SIAM J. Optim. 10(1), 289-313(1999)

5 Huang, X.X, Yang.X.Q, A non -linear Lagrangian approach to constrained optimization problems SIAM J. Optim. 11(4), 1119-1144(2001)

6 Yang.X.Q. , Smoothing Approximation to non- smooth optimization problems . J. Aust. Math. Soc. B 36 , 274-285(1994).

7 Yang, X.Q. , Meng, Z.Q. , Huang, X.X , Pong, G.T.Y., Smoothing non-linear penalty functions for constrained optimization . Numer Function Analysis Optimization 24 (3-4), 351-364(2003)

8 Meng, K.W. , Li, S.J. , Yang, X.Q., A robust SQP method based on a smoothing lower order penalty function optimization 58(1), 22-38(2009).

9 Wu, Z.Y. , Bai, F.S. , Yang, X.Q., Zhang, L.S., An exact lower order penalty function and its smoothing in non-linear programming optimization 53(1), 51-68(2004).

10 Meng, Z.Q., Xu, X.S., Sun, J.W., Huang, L.G., and Shen, R. A second order smooth penalty function algorithm for constrained optimization problem , Computational Optimization and Application , 55,155-172(2013).

11 Wu, Z.Y., Lee, H.W.J., BAi, F.S., Zhang, L.S., Yang, X.M., Quadratic smoothing approximation to l_1 exact penalty function in global optimization . J. Industrial Management Operations 1(4), 533-547(2005)

12 Meng, Z.Q.,Dang, C.Y., Yang, X.Q., On the smoothing of the square root exact penalty function for inequality constrained optimization , Computational Optimization and Application , 35,375-398(2006)

13 Lian,S.J., Smoothing of the lower order exact penalty function for inequality constrained optimization , Journal of Inequalities and applications , Article no 185 (2016)

14 Liu, B.Z., On smoothing exact penalty function for non –linear constrained optimization problems , Journal for Applied Mathematics and computing , 30, 259-270(2009)

15 Lian,S.J, Smoothing approximation to l_1 exact penalty function for inequality constrained optimization , Applied Mathematics and Computation ,219, 3113-3121

16 Pinar, M.C., Zenios,S.A., On smoothing exact penalty functions for convex constrained optimization , SIAM Journal on Optimization , 4, 486-511